10-2004

Phomadidacte: A Computer-Aided Training Program for the Severity Assessment of Phoma Stem Canker of Oilseed Rape

Jean-Noel Aubertot
Unité mixte de recherche d'Agronomie

Yassine Sohbi
Unité mixte de recherche d'Agronomie

Hortense Brun
Unité mixte de recherche d'Agronomie

Annette Penaud
Centre Technique Interprofessionnel des Oléagineux Métropolitains

Forrest W. Nutter Jr.
Iowa State University, fwn@iastate.edu

Integrated Control in Oilseed Crops
IOBC/wprs Bulletin Vol. 29(7) 2006
pp. 247-254

Phomadidacte: A computer-aided training program for the severity assessment of phoma stem canker of oilseed rape

Jean-Noel Aubertot [1], Yassine Sohbi [1], Hortense Brun [2], Annette Penaud [3] and Forrest W. Nutter, Jr. [4]

[1] UMR d'Agronomie INRA/INA P-G, BP1, 78850 Thiverval-Grignon, France; [2] UMR BIO3P, Domaine de la Motte, 35653, Le Rheu Cedex, France; [3] CETIOM, BP4, 78850, Thiverval-Grignon, France; [4] Iowa State University, Dep. Plant Pathology, 351 Bessey Hall, Ames, IA 50011, USA

Abstract: Phoma stem canker (caused by *Leptosphaeria maculans*) is one of the main diseases that affect oilseed rape world-wide. The disease is usually characterised by the visual assessment of the severity of cankers created by the pathogen at the crown level of infected plants. In order to avoid arbitrary categories and to maintain standardisation of assessment keys, a rating scale based on the percentage severity estimates of cross-section cankered crowns has been recently proposed in France. This scale consists in 6 severity classes, defined as a function of the percentage of the discoloured cross-section: 1, healthy plant, no visible lesions; 2, 0-25%; 3, 25-50%; 4, 50-75%; 5, 75-100% of discoloured section; 6, section without any living tissue, plant lodged or broken at the crown level during sampling. However, like many other rating systems for the severity of diseases, assessor bias effects have been reported. The aim of this paper is first to give a description of Phomadidacte, a computer-aided training program to guide assessors on how to use this rating scale; and second, to report an evaluation of its efficacy in training assessors. The basic principle of the program consists of displaying pictures of cankered cross-sections of oilseed rape that are to be rated by the user. These pictures had been previously rated by a panel of eight experts from four different research or extension units (2 from INRA, 1 from CETIOM, and 1 from GEVES), and the mean of the severity classes assigned to each picture were taken as the "true" or correct severity class for each picture. At the end of a training session, the program will either suggest the assessor continues with training or will declare that there was good agreement between the experts and the user, based on two evaluation thresholds: the percentage of pictures correctly rated (75%) and the percentage of pictures for which the severity class given by the user and the expert differ by more than one severity class (5%). A graph summarising the differences between the user and the experts' grades is displayed and an ASCII file containing the data of the training session is created. The user can then browse all the pictures to compare the grades that the user gave with the experts' grades. Phomadidacte runs under Windows® and an on-line version is available. An experiment was conducted to test the efficacy of Phomadidacte in improving assessor accuracy. This experiment compared the grades given by two groups of ten assessors who have been evaluated either with or without using Phomadidacte with the grades assigned by a panel of three experts (INRA, CETIOM, GEVES) on actual diseased field samples. The group evaluated without Phomadidacte used one picture per severity grade and diagrams illustrating the diversity of symptoms that can be encountered. Plants were first graded by experts who chose twenty plants for each severity grade. The plants were ordered randomly from the field and then independently rated by all assessors. The distribution of errors (experts minus the assessors' grade) was significantly different between the two groups of assessors (Kolmogorov-Smirnov test, $P<10^{-4}$). The assessments for the group that had not received Phomadidacte training was in agreement with the experts in 62% of the cases, whereas the group trained with Phomadidacte achieved 69% agreement. Three assessors trained using Phomadidacte succeeded in having a percentage of agreement with experts greater than 75%, whereas none of the non-Phomadidacte group assessors succeeded in achieving this threshold. No assessor, within the two groups, had more than 4.2% of the cases with a difference with the experts that was greater than one severity class, which indicates that the scale is quite easy to use. The generic method described in this paper could be successfully applied to other pathosystems that are difficult to visually characterise.

Key words: Disease severity assessment, phytopathometry, *Leptosphaeria maculans*, blackleg, *Brassica napus*.

Introduction

The assessment of severity due to pest injuries (pathogens, weeds and animal pests) is critical for agriculture. For diseases, there are several reasons for assessing symptoms: to measure the efficacy of different types of control measures (*e.g.* cultivar resistance, chemical, biological, physical or cultural control); to determine damage functions, *i.e.* relationships between injuries and yield losses (that can be used to establish economic thresholds for pesticide applications); to perform regional agronomic diagnoses of commercial fields, such as those defined by Doré et al. (1997); to carry out epidemiological studies on pathogens; to analyse host-pathogen interactions. Several studies have been carried out on direct and indirect methods for disease assessment, and related sampling methods for different crops (Cooke, 1998). In the case of phoma stem canker (*Leptosphaeria maculans)*, one of the most severe diseases of oilseed rape world-wide (West *et al.*, 2001), several studies have been conducted to improve the characterisation of disease injury (Gilligan 1980; Pierre and Regnault, 1982; Rimmer and Van den Berg, 1992; Van den Berg *et al.*, 1993; Rempel and Hall, 1996; Aubertot *et al*., 2004). However, unlike other pathosystems for which computer-aided training programs have been developed – *e.g.* DisTrain (Tomerlin and Howell, 1988); Disease.Pro (Nutter and Schultz, 1995); WinCombro (Canteri and Giglioti, 1998) – little efforts have been made to develop methods to help train assessors to assess the severity of phoma stem canker.

The basic principle of these computer-aided training programs consists of: i) displaying computer generated images of infected leaves whose visible diseased areas are known precisely; ii) letting the user rate these images; iii) comparing the user's rating with the computer generated reference. After each evaluation, these computer-aided training programs can be evaluated to determine if they improve the accuracy of assessors in estimating the percentage disease severity on a sampling unit (leaf). However, the pathosystems considered in these types of computer-aided training programs are foliar-spotting (*e.g.*, leaf rust, powdery mildew, septoria, scald, spot blotch, net blotch, leaf rust, stem rust on wheat; rust, early leaf spot, late leaf spot on peanut). In these cases, the computer generated references consist of symptoms represented with one or two colours positioned on monochromatic leaves (green) or pictures of leaves. In the case of phoma stem canker, the symptoms to assess are discoloured tissues of stem cross-sections that show a wide range of shades, so it is very difficult to generate reference computer images. In addition, field observations sometime present perturbing elements on the observed cross-sections: larva injuries, small soil aggregates, *etc*. The objective of this paper was i) to present a method to overcome these difficulties; ii) to describe Phomadidacte, a computer computer-aided training program for the severity assessment of phoma stem canker of oilseed rape; and iii) to present an evaluation of the efficacy of Phomadidacte to train assessors.

Materials and methods

Construction of visual references without computer generated images

In order to take into account the complexity of symptomatic cross sections to assess, one can consider using pictures to represent various injury classes to develop clear categories for reference. One way of producing such references categories would be to perform an

automated image analysis of pictures of symptoms, to assess, for instance, the proportion of diseased tissues. However, practice shows that such an approach is inappropriate because of the difficulty to define proper grey-level thresholds to distinguish diseased areas (pixels) from healthy tissues. These thresholds can vary from picture to picture, or even within a given picture. This is why we proposed to develop a set of disease reference images using the expertise of experienced assessors. The basic principal of the method consists of independent experts rating a set of pictures representing a wide range of disease severities, according to a given rating scale. If the majority of experts agree as to the grade of a specific picture, this picture is integrated into the reference image database. As with other computer-aided programs, this reference can subsequently be used to train assessors. One can remark that other authors have already accepted the rating provided by the developer of a rating scale to be the "true" or "actual" values for the disease severity of a sampling unit (O'Brien and van Bruggen, 1992).

Development of a reference for Phomadidacte

Phoma stem canker is usually characterised by the visual assessment of the severity of cankers created by the pathogen at the crown level of the plants. In order to avoid arbitrary categories and to maintain standardisation of assessment keys, a rating scale based on the percentage severity estimates of cross-sections of cankered crowns has been recently proposed (Aubertot *et al*., 2004). This scale consists of six severity classes defined as a function of the percentage of the discoloured cross-section: 1, healthy plant, no visible lesions; 2, 0-25% of discoloured cross-section; 3, 25-50% of discoloured cross-section; 4, 50-75% of discoloured cross-section; 5, 75-100% of discoloured section; 6, section without any living tissue, plant lodged or broken at the crown level during sampling (Aubertot *et al*., 2004). A set of pictures of more or less cankered cross-sections of oilseed rape stems was created in July 2002. These have been rated by a panel of eight experts from four different research or extension units (2 from INRA, 1 from CETIOM, and 1 from GEVES). This rating scale was used to build a reference image database of 120 pictures (20 pictures per severity class).

Structure of Phomadidacte

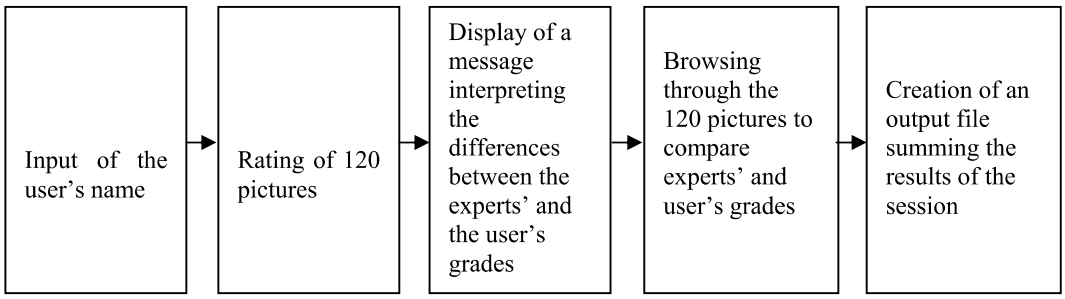

Figure 1. Process flow chart of Phomadidacte, a computer-aided training program for the severity assessment of phoma stem canker of oilseed rape caused by *Leptosphaeria maculans*.

The structure of Phomadidacte is presented Figure 1. After each assessor has specified their name, the user assigns each of the 120 references pictures (that appear in a random order) to one of the six severity classes (Figure 2). After all 120 disease images have been

graded, the distribution of assessor error (*i.e.*, the differences between the experts' grades and the corresponding assessor's grades) is displayed along with a short message that interprets the results. The message either congratulates the assessor for being in good agreement with the experts, or suggests the assessor (user) continue training before assessing Phoma stem canker severity in actual field experiments. The content of the message depends on two evaluation criteria: the proportion of pictures correctly graded and the proportion of pictures with a difference (from the expert's) greater than one severity class. In order to be congratulated, the percentage of pictures correctly rated has to reach at least 75% and the percentage of pictures for which the severity class given by the user and the experts differ by more than one severity grade has to be lower than 5%. These two evaluation thresholds were defined using the inter-variability of the experts: each expert would satisfactorily fulfil these two conditions when compared with the other 7 experts. A graph summarising the differences between the user and the experts' grades is displayed and an ASCII file containing the training session data is created for subsequent statistical analyses. The user can then browse all 120 pictures to compare the grades that they provided, versus the experts' grades (Figure 2). Phomadidacte operates under Windows® and an on-line version is available at: http://www-agronomie.grignon.inra.fr/fintranet.html. A login and a password will be supplied on request to the corresponding author (aubertot@grignon.inra.fr).

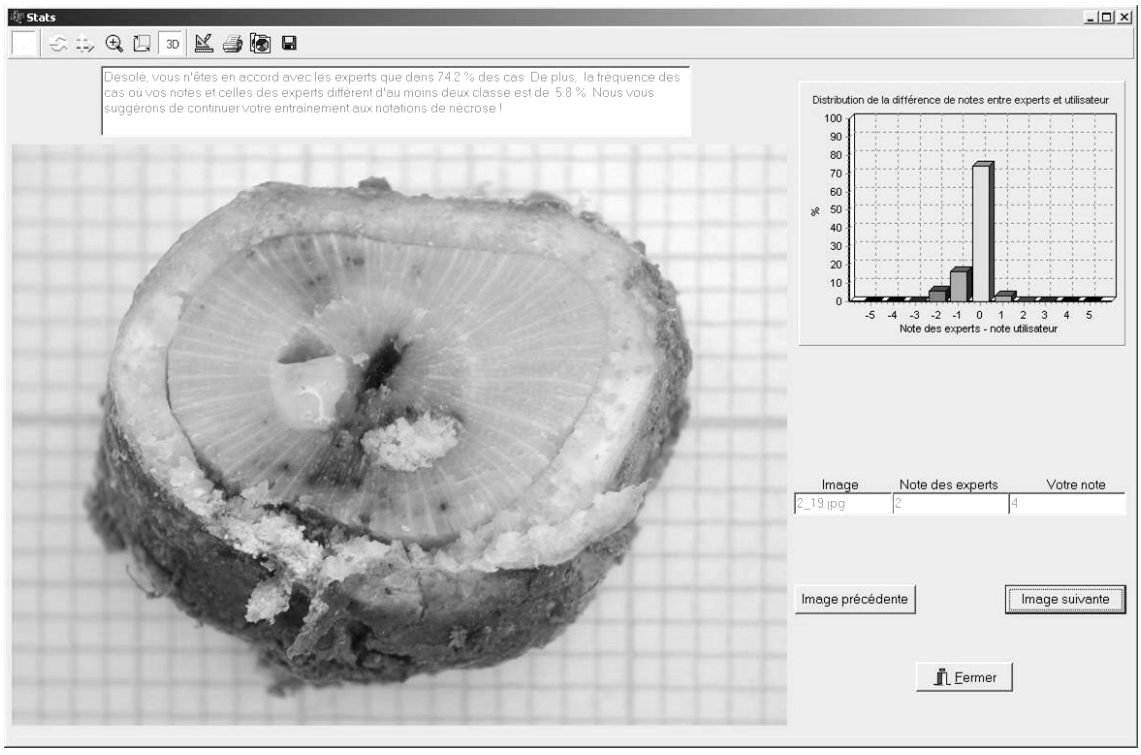

Figure 2. Screen image of Phomadidacte, a computer-aided training program for the severity assessment of phoma stem canker of oilseed rape, at the end of an evaluation and training session. A message indicates whether or not the user is in satisfactory agreement with the experts. A graph summarising the differences between the user and the experts' grades that were assigned for each of the 120 sampling units (pictures) is displayed. The user can then choose to browse all of the pictures to compare the grades that they provided versus those given by the experts' grades.

Evaluation of Phomadidacte

An experiment was done to compare the grades given by two groups of ten assessors who have been trained to assess Phoma stem canker disease of oilseed rape (caused by *Leptosphaeria maculans*) either with or without using Phomadidacte. The "true" or accurate disease severity class for each field sample was determined by a panel of three experts (one each from INRA, CETIOM, and GEVES). None of the twenty assessors had previously assessed the severity of Phoma stem canker on oilseed rape. The group trained without Phomadidacte used one picture per severity grade and diagrams illustrating the diversity of symptoms that might be encountered (Figure 3). Plants were first graded by experts who chose twenty plants for each severity grade. The plants were randomly reordered and then independently assessed by all the twenty assessors. We used SAS Release 6.12 for Windows (SAS Institute Inc., 1989) for statistical analyses. The distribution of errors (experts minus assessor's grades) was tested with the non-parametric Kolmogorov-Smirnov of the NPAR1WAY procedure.

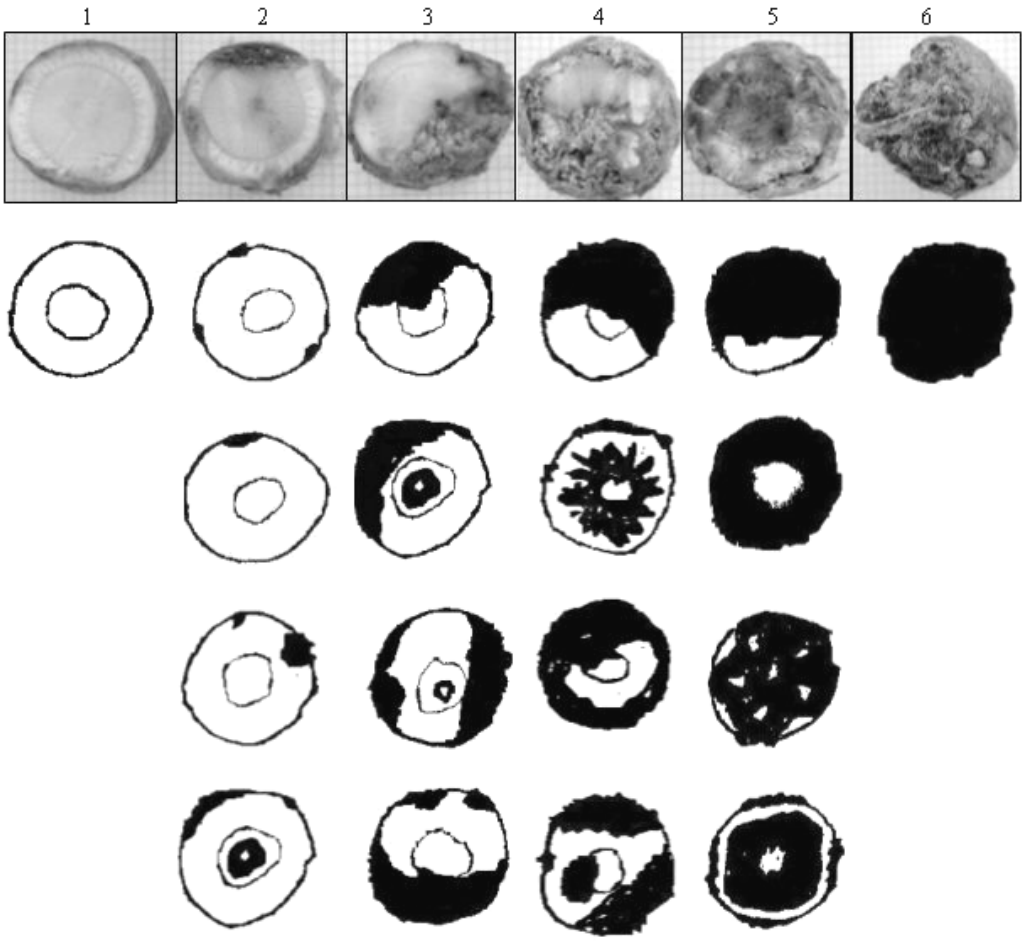

Figure 3. Images and sketches of cankered crown cross-sections of oilseed rape (Phoma stem canker, caused by *Leptosphaeria maculans*) classified into 6 severity classes: 1, healthy plant, no visible lesions; 2, 0-25% of discoloured cross-section; 3, 25-50%; 4, 50-75%; 5, 75-100% of discoloured cross-section; 6, section without any living tissue, plant lodged or broken at the crown level during sampling (Aubertot *et al.*, 2004). The graph paper used is graduated in millimeters.

Results and discussion

The distribution of errors (experts minus assessors' grades) was significantly different between the two groups of assessors (Figure 4; Kolmogorov-Smirnov test, $P<10^{-4}$). No difference greater than two severity grades were observed for both groups. No assessor (within the two groups) had more than 4.2% of the cases with a difference with the experts that was greater than one severity class. This indicates that the scale is quite easy to use, even when assessors are trained with diagrams instead of pictures.

The group trained without Phomadidacte was in agreement with the experts in 62% of the cases, whereas the group trained with Phomadidacte achieved 69% agreement. In addition to this increase of agreement between assessors and experts, Phomadidacte led to a symmetric distribution of errors (Figure 4), whereas the group trained without Phomadidacte overestimated the severity of symptoms as compared to the experts. Three assessors within the Phomadidacte-trained group succeeded in having a percentage of agreement with the experts greater than 75%, whereas none of the assessors in the control group succeeded in achieving this threshold.

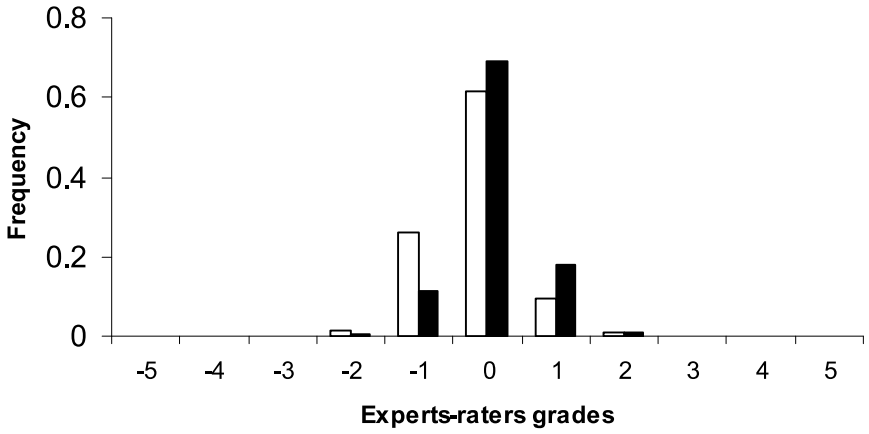

Figure 4. Distribution of errors (experts minus assessor's grades). □: represents the group of ten assessors trained with pictures and diagrams that are presented in Figure 3; ■: represents the group of ten assessors trained with Phomadidacte. Twenty oilseed rape cross-sections of crowns infected with *Leptosphaeria maculans* were used per severity grade were assessed by each group of ten assessors. The actual or "true" severity grades of each reference were estimated on real plants by three confirmed experts.

The main difficulty that is encountered when developing computer-aided training programs for disease severity assessment is the establishment of a "standard" to train or "calibrate" each assessor (Cooke, 1998). Image analysis could appear an objective way to quantify disease severity. However, since a threshold has to be used to distinguish healthy from diseased tissues, image analysis also suffers from a lack of objectiveness (Nutter *et al.*, 1993). The originality of Phomadidacte consists not only in using pictures of symptoms to train assessors, but also in using the expertise of a set of experts to develop a disease severity standard for each disease class. This approach should not be seen as an alternative to

computer-aided training programs that use generated diagrams, but as a complementary method where computer-generated images are not practical. Such programs are aimed to train assessors to assess the percentage of the diseased area on plant organs (leaves, generally) for contrasted objects (*e.g.* orange pustules of brown rust on green leaves of wheat), whereas for pathosystems such as phoma stem canker that express symptoms with a wide range of discolouration, and shapes within a disease severity class would be very difficult to mimic with computer generated pictures. This is why the approach presented in this paper should be of value for various other pathosystems. The drawback of the method is that the standard used is actually based upon the agreement of specialists, and is not really a true measurement of the disease severity. This is why it is important to develop the standard images through the use of a panel of experts with as much experience as possible to ensure the proposed standards are representative.

Acknowledgments

The authors thank D Le Floch and R Gosse for their technical assistance; the eight experts from INRA, CETIOM, and GEVES for having willingly agreed to rate hundreds of pictures of cankered cross-sections; Jean-Philippe Maigniel (GEVES) and the twenty assessor apprentices who kindly accepted the invitation to participate in the evaluation of Phomadidacte.

References

Aubertot, J.N., Schott, J.J., Penaud, A., Brun, H. & Doré, T. 2004: Methods for sampling and assessment in relation to the spatial pattern of phoma stem canker (*Leptosphaeria maculans*) in oilseed rape. – European Journal of Plant Pathology 110: 183-192.

Canteri, M.G. & Giglioti, E.A. 1998: Combro: software for selection and training people to estimate severity of sugarcane rust and borer-rot complex. – Summa Phytopathologica 24: 190-192.

Cooke, B.M. 1998: Disease assessment and yield loss. – In: Jones, D.G. (ed.). The Epidemiology of Plant Disease. Kluwer Academic Publishers, Dordrecht (The Netherlands): 42-72.

Doré, T., Sebillotte, M. & Meynard, J.M. 1997: A diagnostic method for assessing regional variations in crop yield. – Agricultural Systems 54: 169-188.

Gilligan, C.A. 1980: Size and shape of sampling units for estimating incidence of stem canker on oil-seed rape stubble in field plots after swathing. – Journal of Agricultural Science 94: 493-496.

Nutter, F.W., Gleason, M.L., Jenco, J.H. & Christians, N.C. 1993: Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. – Phytopathology 83: 806-812.

Nutter, F.W. & Schultz, P. 1995: Improving the accuracy and precision of disease assessment: selection of methods and use of computer-aided training programs. – Canadian Journal of Plant Pathology 17: 174-184

O'Brien, R.D. & van Bruggen, A.H.C. 1992: Accuracy, precision, and correlation to yield loss of disease severity scales for corky root of lettuce. – Phytopathology 82: 91-96.

Pierre, J.G. & Regnault, Y. 1982: Contribution à la mise au point d'une méthode de plein champ destinée à mesurer la sensibilité des variétés de colza au phoma. – Informations Techniques du CETIOM 81(IV): 3-18.

Rempel, C.B. & Hall, R. 1996: Comparison of disease measures for assessing resistance in canola (*Brassica napus*) to blackleg (*Leptosphaeria maculans*). – Canadian Journal of Botany 74: 1930-1936.

Rimmer, S.R. & van den Berg, C.G.J. 1992: Resistance of oilseed brassica ssp. to blackleg caused by *Leptosphaeria maculans*. – Canadian Journal of Plant Pathology 14: 56-66.

SAS Institute. 1989: SAS/STAT User's Guide. Version 6. – SAS Institute, Cary, NC.

Tomerlin, J.R. & Howell, T.A. 1988: DISTRAIN: a computer program for training people to estimate disease severity on cereal leaves. – Plant Disease 72: 455-459.

van den Berg, C.G.J., Rimmer, S.R. & Parks, P. 1993: Comparison of scales for rating severity of blackleg in canola. – Canadian Journal of Plant Pathology 15: 49-53.

West, J.S., Kharbanda, P.D., Barbetti, M.J. & Fitt, B.D.L. 2001: Epidemiology and management of *Leptosphaeria maculans* (phoma stem canker) on oilseed rape in Australia, Canada and Europe. – Plant Pathology 50: 10-27.